

When minoritized languages encounter MT: perceptions and expectations of the Basque community

Nora Aranberri, HiTZ Center, University of the Basque Country (UPV/EHU)
Uxoia Iñurrieta, GOI Institute, Basque Summer University (UEU)

ABSTRACT

Machine translation (MT) is improving even for low-resource minoritized languages such as Basque, for which free online engines are available. However, the level of adoption and common practices involving the technology are unknown, even though it has the potential to disrupt a carefully planned Basque language revitalization and sustainability process. To shed light on MT usage habits and perceptions among the Basque community, we report on the results of a survey of language specialists and general users, and a focus group with professional translators and interpreters. The data shows that MT is already becoming more popular among users of all backgrounds and that, overall, the attitude towards the technology is positive, which might result in increased use in the future. However, participants express concern about the impact MT will have on the development of Basque. The results call for further research on the language impact of MT and MT literacy initiatives.

KEYWORDS

machine translation, minoritized languages, Basque, perceptions, sustainable revitalization, professional translators, users.

1. Introduction

Machine translation (MT) research has so far mainly focused on improving the technical capabilities of the technology, and the study of its interaction with users has been mostly limited to analysing how it performs within professional translation settings. However, with more and more powerful systems freely available online, translators are not the only users of the technology. People with varying language proficiencies in their working languages also turn to MT to address their diverse translation needs.

It is not unreasonable to think that the use of this tool will cause changes in the ways people interact with each other. And yet, the interaction between MT and general users is significantly under-researched. The works of Nurminen and Papula (2018) and Vieira *et al.* (2023) are two attempts at uncovering usage patterns, but we still have a long way to go to fully understand who is using MT, how, and for what purposes.

It is important to keep in mind that engines do not produce perfect translations and that MT and self-post-editing do not seem to produce texts with the exact same characteristics as those directly written in the target language (Aranberri 2020, De Clercq *et al.* 2021, Vanmassenhove *et al.* 2021). It is also not far-fetched to think that, given the continuous use of the technology, *MT-language* might influence target languages (Sánchez-Gijón and Piqué Huerta, 2020). This is true for all languages, but it is particularly relevant for minority languages and those in a normalization

process, such as Basque. All these factors raise the question of the extent to which people are aware of the benefits, risks, and limitations of MT, leading to a call for initiatives in MT literacy (Bowker and Buitrago 2019, Bowker 2020, Kenny 2022).

In this article, we aim to shed some light on the role MT plays within the Basque community. To that end, we start by providing a short description of the linguistic context of the Basque language. We then report on the results of a survey of MT habits and opinions of both language professionals and general users. Finally, we summarize the thoughts and ideas of professional translators regarding MT and post-editing. Overall, we conclude that MT for Basque is a reality, and that people of all backgrounds use it, which will most probably impact the use of the language. Results highlight the need for research on MT from a sociolinguistic perspective.

2. The Basque context

The considerable improvement in quality that MT has achieved in recent years has resulted in its steady adoption as a language tool. MT is perceived as a tool that facilitates the comprehension of texts written in foreign languages but also for the production of texts in general: it can help in the drafting of bilingual documents and texts in languages that the user has not mastered. In short, today, MT engines are being made available to all types of users to deal with an increasingly wide range of languages and increasingly diverse multilingual contexts.

This is precisely the case of the social context on which the present article focuses: the Basque-speaking community. Basque is a minoritized language whose use is spread between two regions in Spain and one in France: it is an official language in the Basque Autonomous Community (BAC) and in part of the Autonomous Community of Navarre in Spain, but it has no official status in the region of New Aquitaine in France. As such, it coexists with two hegemonic languages: Spanish and French.

According to the 2021 Population and Housing Census of the BAC — where 85% of the Basque speakers reside — 62.4% of the population aged 2 and over, namely, 1,349,808 people, have some knowledge of Basque (Eustat, 2022). Based on their level of knowledge, 936,812 are people who can understand and speak the language well, and 412,996 are quasi-Basque speakers, that is, people with a good or regular level of understanding, but with difficulty in production. A search in the Data Bank of the Basque Institute of Statistics (Eustat 2023) shows that only 40 years earlier, in 1981, the population with knowledge of Basque was around 36%. This current situation is a clear result of a revitalization process which started in the 1960s with various social initiatives (such as the *Ikastolas* or Basque schools) and was then followed by another crucial step: the region's self-government through the 1979 Statute of Autonomy.

In Baztarrika's view (2019), the revitalization of Basque in the BAC is an example of a sustainable process, enabled only by a coordinated effort made at multiple levels. It all started with a shared view of what the language had to become. Setting other models aside (consider the approaches in Belgium, Canada, and France, for example), the Basque community sought the complementarity and balance of both languages, Basque and Spanish, rather than perpetuating two linguistic communities with a hierarchization of the languages.

Needless to say, the capacity for political decision-making acquired by the local government at the time played a key role in the execution of the aforementioned model. Above all, it allowed the passing of the Law for the Normalization of the Basque Language (Law 10/1982), resulting in policies being put in place to invest in several vital areas. Education was one such area, with the 1980s seeing studies at all levels offered in Basque for the first time. Media was also among the first sectors to be promoted, and Basque radio and television stations were launched in 1982 and 1983, respectively. Importantly, the local public administration started a Basque alphabetization process as well, to ensure the citizens' right to interact with it in Basque.

While there is no denying that the revitalization effort of the past decades has increased the knowledge and use of Basque, there has not yet been full normalization of the language. It is, however, a challenge that many administrative and social players aim to respond to, as sustaining and developing multilingualism has been identified as a factor of progress, social cohesion and community strengthening (Baztarrika 2019: 242).

If we focus on the digital realm, the report by Gurrutxa and Ceberio (2017) states that Basque is regularly used on the internet, even for social media. The authors claim, however, that although localized digital services and interfaces exist in Basque, users often opt for the versions in major languages. This trend has been revealed by several studies and pinpoints the higher maturity and quality of the technology in major languages as the main reason. Studies seem to indicate that using minority languages in this context requires "a good amount of perseverance, will, and resilience" (Soria 2023: n.p.). A key issue that we must bear in mind is that, being bilingual, Basque society is digitally advanced, that is, even when certain information and accurate applications are not available in Basque, access to them is guaranteed via Spanish and French. Therefore, using Basque is most often a personal choice — or a need that arises in the workplace. This means that Basque is in constant competition with the high-resource languages by which it is surrounded.

In this context, we could posit that MT can serve as a tool to promote such sustainable development of minoritized languages for several reasons: it is available at our fingers, it can deal with large volumes of text at high speeds if necessary, and its quality is ever increasing. Regarding Basque, in less

than a year, during 2020, three neural engines joined Google Translate in offering automated translations for the Spanish-Basque language pair: namely Elia, Batua, and itzuli. The engines, which currently also include French and English, Catalan, Galician and the Biscayan dialect, have been made publicly available under the assumption that they can provide a useful level of translation quality. Admittedly, MT engines for Basque tend to lag behind those for major languages, mainly due to the lack of training data. In addition, the fact that Basque is an isolated, agglutinative, and free word-ordering language — in other words, distant to the languages it is usually paired with — tends to play a detrimental role in the training of the systems. Yet, it is undeniable that the technology is rapidly advancing and available to all.

Nevertheless, we should not ignore the potential benefits of the technology, as its influence on the target language and on communication in general is still understudied. Exploratory research suggests that translations produced by MT show reduced lexical and morphological richness, with common language patterns reused even more frequently and rarer patterns lost (Vanmassenhove *et al.* 2021). On top of that, different types of architectures, supervised and unsupervised systems, for example, might produce distinct language styles regarding structural preferences (Marchisio, Freitag and Grangier 2022). Initial research about linguistic competence and MT also shows that, given an intermediate level in a foreign language, it might be more effective from an expressive perspective to start by writing a text in one's mother tongue before using MT to produce the foreign language version (Aranberri 2020).

Therefore, while new solid conclusions are still to be drawn on the impact of MT on target languages, we should take steps to partner MT with sustainable language development. An opportunity for this emerges along the lines of the MT literacy initiatives proposed by Bowker and Buitrago Ciro. In their words, MT literacy can be defined as “users’ capacity to understand how MT systems work and can be used, to evaluate the MT-friendliness of a text, and to modify MT output” (Bowker and Buitrago Ciro, 2019: 88). In this way, the MT Literacy Project led by Bowker sets the foundations for training initiatives directed at general citizens. To maximize effectiveness, users should be trained in

“[...] deciding whether MT is the right tool for a job, evaluating whether information is too sensitive to paste into a free online tool, learning how to write in a translation-friendly way, recognizing common (and not-so-common) MT errors, and learning how to fix them.” (The MT Literacy Project n.d.)

It goes without saying that the depth of the five aspects listed by Bowker's project will vary according to the sociolinguistic contexts and the features of the user groups. Even information about who uses MT, how and for what purpose, is unknown for Basque. Yet, the potential of the tool is evident. Although the figures for Basque knowledge are encouraging, we must not

forget the other side of the coin: 37.6% of the population do not speak Basque and 19% are not able to fully understand or competently express themselves using it. Adding to this, according to the latest Sociolinguistic Survey (Eustat 2016), around 29% of Basque speakers (12% of the total population) have a preference for Spanish over Basque. Therefore, it is likely that, given a situation where a bilingual Basque-Spanish text must be produced, around 68% of the population would write it in Spanish and then use an MT engine to obtain the Basque version. It would be reasonable to assume that at least this portion of the citizenry would have difficulty, to a greater or lesser degree, in evaluating and producing texts in correct, accurate and natural Basque. This highlights the central role — positive and negative — MT could play in this scenario, which should not be overlooked by the agents of the revitalization process.

Research suggests that citizens of different languages are adopting the technology (Nurminen and Papula, 2018; Vieira *et al.* 2023) and that the positive portrayal by the media, which tends to omit the implications of its use (Vieira, 2020), might encourage this trend. The more fluent language present in the translations of current MT systems could also play a role, as output comprehensibility seems to be a key factor in the degree to which people trust MT (Rossetti *et al.* 2020). As shown by Delormes Benites (2021), unguarded use could even have discourse-related consequences. Focusing on Switzerland, the author found that the topics that reach a target-language community are not necessarily the same as when the news was originally written prior to being (machine-)translated. This highlights the risk of using translated texts as a sole mode to inform a speaker community, as texts tend to be directed at the source language speaker community. In the case of Irish, a minority language, a survey of Irish translators also revealed conflicting perceptions regarding MT (Moorkens 2020). In contrast to government-employed translators, freelancers showed concern over working conditions, decreasing translation quality and even translation competence, and in general, they emphasized the importance of language policies for minority languages.

As a first step in defining how Basque society could be informed about MT use, in this work, we aim to get an insight into how people view and approach Basque MT.

3. Survey on MT practices and views

We conducted a survey on MT practices and views to explore the status of MT adoption within the Basque society and for Basque. Considering the implications of MT use for the minoritized language, it was our aim to learn about the technology's perceived influence on translation flows. In this section, we first describe the question blocks included in the survey, the process of distribution, and the respondent profile, and next we reflect on the results.

3.1. Survey questions

We divided the survey into three blocks. The first focused on translation practices in general. It was deemed important to get an overview of translation needs and habits of the speakers to better understand the role MT can play in such contexts. Therefore, a set of questions was formulated to learn about where and what they translated and which languages were involved in the process. We next turned to MT. We reused previous questions, this time directed at MT, and added several new ones to learn about how they interact with the technology. The final block focused on the expected impact of MT on Basque.

3.2. Survey dissemination

We created a full survey which included information about its aim and what it involved so that participants could provide their consent. Then followed the three sets of questions.

The survey was constructed using Google Forms, originally in Basque and then translated into Spanish and French. The links to the three versions were distributed in all communications for the participants to choose the language with which they felt more comfortable.

The survey was disseminated through numerous channels, which included the electronic bulletin board of the University of the Basque Country, the distribution list of EIZIE (the Association of Translators, Correctors and Interpreters of Basque Language), WhatsApp and Telegram groups, and targeted emails to translation companies, among others.

3.3. Survey participants

Over a period of two months, we received a total of 1143 responses (801, 330 and 12 in the Basque, Spanish and French versions respectively). We collected participant information regarding gender, age, language proficiency and specialization to help us outline their profile.

61.15% of respondents are female, 38.06% are male and 0.79% non-binary. The difference is rather evenly distributed across age groups, except for the 56-65 range, where the number of male and female respondents is very similar (see Table 1).

With regard to age groups, we were able to collect a good number of responses for all ranges (336 for the 18-25 group, 168 for the 26-35 group, 245 for the 36-45 group, 208 for the 46-55 group, and 154 for the 56-65 group), except for the below 18 and above 65 categories, which consist of 16 responses each (see Table 1).

Age groups	<18	18-25	26-35	36-45	46-55	56-65	>65	TOTAL
Female	1.14	18.64	9.80	13.56	10.76	7.00	0.26	61.15
Male	0.26	10.50	4.81	7.70	7.35	6.30	1.14	38.06
Non-binary	0.00	0.26	0.09	0.17	0.09	0.17	0.00	0.79
TOTAL	1.40	29.40	14.70	21.43	18.20	13.47	1.40	100

Table 1: Percentage of respondents per gender group across age groups.

In reference to Basque proficiency, few respondents have an elementary or no command of Basque (6.04% and 3.67%, respectively), and a slightly higher number report an intermediate proficiency (12.95%) (see Table 2). About three respondents in four, that is, 77.34%, report having advanced knowledge. This reveals that our respondent profile is skewed towards speakers who are skilled in Basque.

We observed that Basque proficiency follows the same tendency for all genders (see Table 2) and age groups (see Table 3), although a couple of exceptions can be spotted, which reflect the revitalization process. Only half of the youngest participants report an advanced level of command, while 37.50% – 20 points higher than other categories – report an intermediate level. However, this is expected for young learners. Within the Basque educational system, secondary education students are expected to achieve a B2 level of the language of instruction. It is usually during the final high school year and those following that young people prepare to sit the C1 examination. In fact, the data displays this behaviour to a certain extent, as we see a decrease in the intermediate level and an increase in the advanced level for the 26-35 and 36-45 categories.

	None	Elementary (A1-A2)	Intermediate (B1-B2)	Advanced (C1-C2)	Total % of respondents
Female	3.15	6.04	12.95	77.83	61.15
Male	4.60	5.29	13.73	76.32	38.06
Non-binary	0.00	1.11	11.95	88.89	0.79
Total % of reported Basque competence	3.67	6.04	12.95	77.34	

Table 2: Percentage of participants per gender across levels of competence in Basque.

	None	Elementary (A1-A2)	Intermediate (B1-B2)	Advanced (C1-C2)	Total % of respondents
<18	0.00	12.50	37.50	50.00	1.40
18-25	4.46	7.74	16.67	71.13	29.40
26-35	5.36	5.95	5.95	82.74	14.70
36-45	1.22	3.27	7.76	87.76	21.43
46-55	1.92	4.81	12.50	80.77	18.20
56-65	6.49	8.44	16.88	68.18	13.47
>65	6.25	0.00	31.25	62.50	1.40
Total % of reported Basque competence	3.67	6.04	12.95	77.34	

Table 3: Percentage of respondents per age group across levels of competence in Basque.

A feature that might affect MT practices and perceptions is the level of language specialization of the respondents, that is, whether they work or study in the field of languages. We hypothesise that those working in language-related areas might be more sensitive to quality in translations and the language MT produces, and more conscious about its use, while those who study or are professionals in other fields will most probably view translation and MT as mere gisting tools for communication, and they might not give too much consideration to precision and fluency issues. The data reveals that 64.13% of the respondents do not belong to the language specialist group (a.k.a. other users), while the remaining 35.87% does (see Table 4). Out of the latter, 35.85% are translation and/or interpreting professionals or students of the discipline (a.k.a. translators and interpreters), and 64.15% work in other language-related areas (a.k.a. language professionals). Given that a considerable number of responses were collected for each of the three groups, results will be provided separately to monitor any divergences.

Language specialization	Absolute numbers	Percentage
Translators and interpreters	147	12.86 %
Language professionals	263	23.01 %
Other users	733	64.13 %

Table 4: Number of respondents across language specialization groups.

Overall, we can say that the collected responses cover the majority of gender and age groups rather acceptably. Also, we will be able to compare practices and opinions from both language specialists (including translators

and interpreters) and other users. Our sample is somewhat unbalanced towards advanced speakers of the language, and this is something we should bear in mind when generalizing results.

3.4. Results I: translation practices

Let us first look at the general translation practice of respondents. This should provide us with an insight into the overall needs and efforts, some of which might be aided or replaced by MT in the future.

When asked about the frequency with which respondents engage in translation, few indicate that they are never faced with this task (see Figure 1). Interestingly, even out of those who are not involved in language-related disciplines (other users) only 3% report not having translation needs – in fact, these users report high translation activity. 38.61% report addressing translations sometimes or very often. In turn, not surprisingly, almost 80% of translators and interpreters report translating very often. For language specialists, the frequency decreases, 80% admit to translating sometimes or very often. When specifically asked about Basque, rather evenly, the three groups report Basque as being a major part of their translation efforts, with an average of 85.65% across groups (see Figure 1). It could be argued that respondents encounter translation needs rather frequently, most often involving Basque.

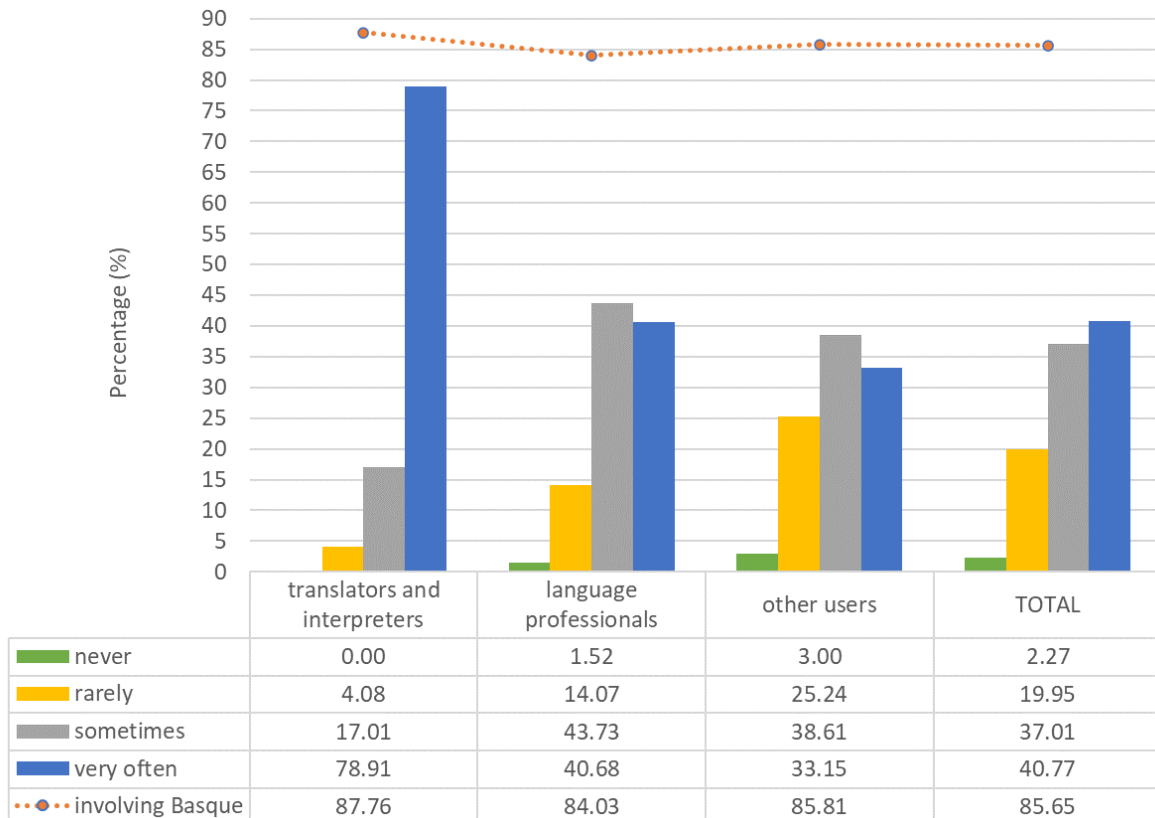


Figure 1: Do you engage or have you ever engaged in translation in your everyday life? That is, do you find yourself having to say/write what you read/listen to in a language in a different one?

To pinpoint the communicative situations where translation from or into Basque arises, we asked respondents to state the extent to which they engaged in translation within three main scenarios: the workplace, educational contexts, and the private sphere. The answers of the respondents who reported to having to address Basque translation for whom the respective scenarios are applicable are displayed in Figure 2.

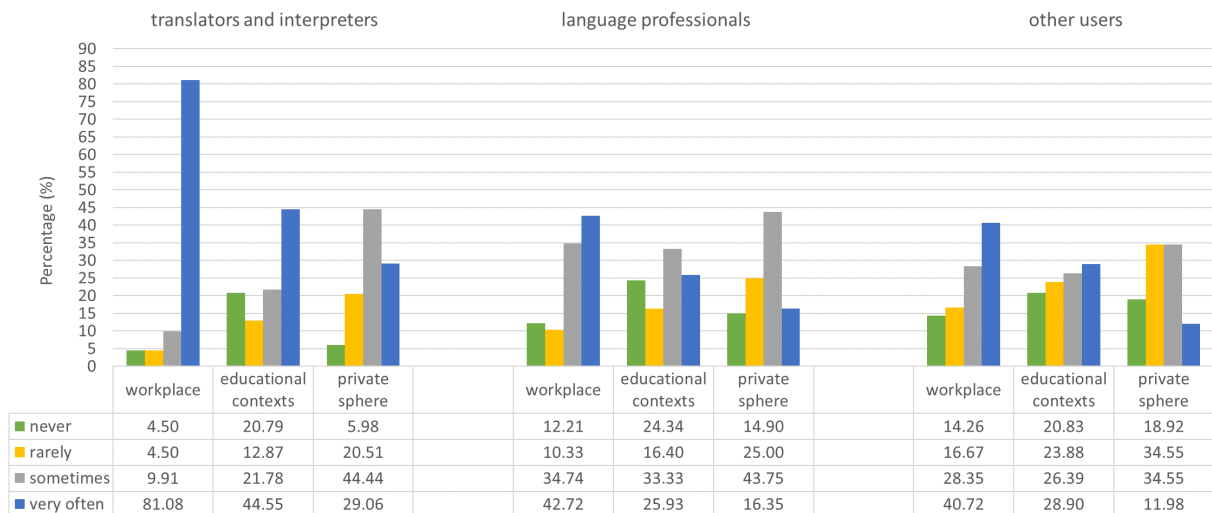


Figure 2: In which contexts do you engage in translation? Select the frequency.

As expected, we observe that translators and interpreters handle Basque translation in the workplace very often (81.80%). Only 4.5% never do so. What is interesting is that respondents who do not belong to this industry also report involvement in translation from/into Basque in the workplace. Language professionals and other users claim to engage in translation involving Basque very often (42.72% and 40.72%, respectively). The percentage increases to around 70% if we also consider those admitting to engaging in this activity sometimes.

Educational contexts also see a rather significant percentage of respondents requiring translation from/into Basque very often (25.93% and 28.90% of language professionals and other users, respectively), while the figures decrease in the private sphere (16.35% and 11.98%, respectively). A possible reason for this is that the use of Basque might be compulsory in many educational settings and therefore learners must deal with Basque texts, whereas engagement in personal interactions in Basque is freer. It is interesting to see that the higher the level of specialization in languages, the higher the involvement in Basque translation is outside the workplace.

A final question in this block refers to the types of texts respondents translate. We asked participants to select the text genres that applied to them from a list. Results indicate that there are differences between user profiles (see Figure 3). The most salient difference is the higher proportion with which other users translate emails and other messages and essays and

other writings for learning purposes compared to professional translators and interpreters. We might have expected that other users would engage in translation for leisure and entertainment more often than in other genres, but this does not emerge from their answers. This seems to be in line with the prominence of using translation in the workplace. Professional translators address literature and audiovisual texts more frequently than other groups and mostly work on reports and administrative texts. For language professionals, messages and informative texts are key. Finally, it is interesting to see the relevance of oral texts across groups.

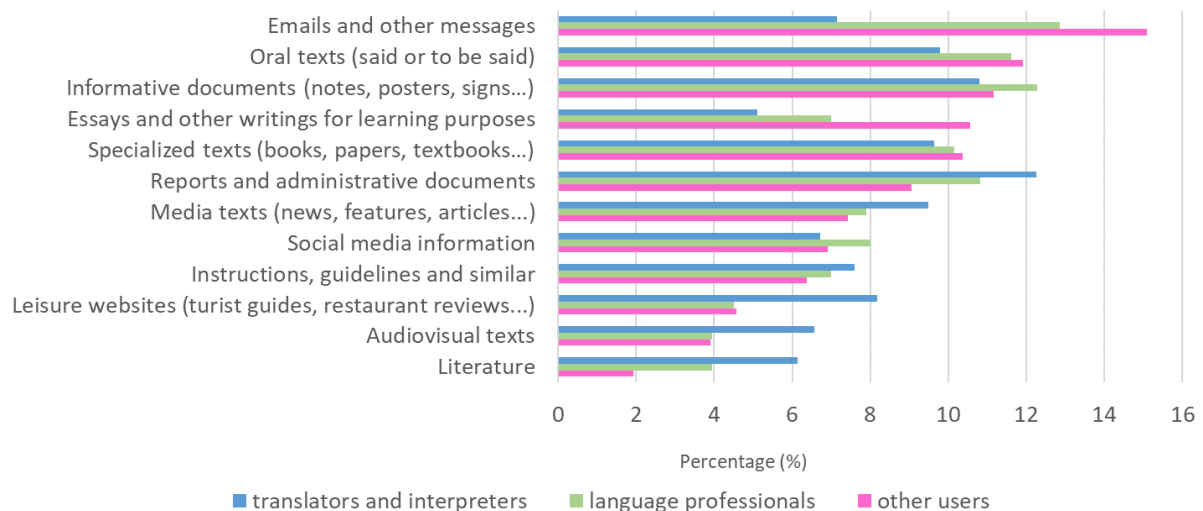


Figure 3: What do you translate? Select all that apply.

3.5. Results II: machine translation practices

In reference to MT use, 26.25% of all respondents admit to using the technology very often, the percentage being very similar across groups (see Figure 4). The proportion of respondents who use it sometimes increases up to 42-43% in the case of language professionals and other users, and it is slightly lower at 35.37% for translators and interpreters. Only 7.61% of the total respondents report never availing of MT, again with no differences across groups. This seems to indicate that the adoption of the technology is widespread. Interestingly, 72.18% of all respondents acknowledge using MT for Basque too, with hardly any difference across groups. Therefore, we see that despite the engines for this language lagging in terms of quality, their use is spreading.

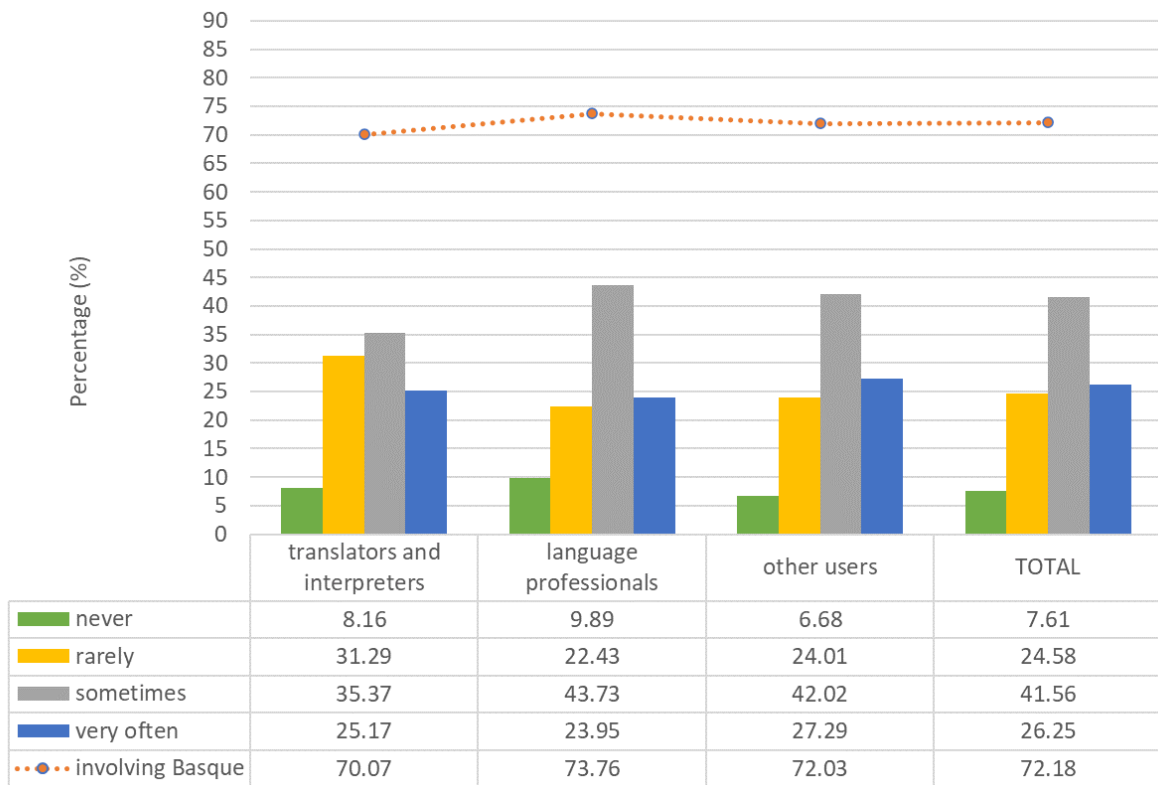


Figure 4: Do you use MT to translate from or into Basque?

The general purpose for which respondents use MT is revealing (see Figure 5). When asked about the objective of using MT for Basque, the less frequent aim is to (better) understand the language. This is in line with the linguistic proficiency reported by respondents, as the majority should not face serious comprehension issues. Expectedly, we see that it is other users who are more active in this activity: 29.55% admit to using MT for this purpose.

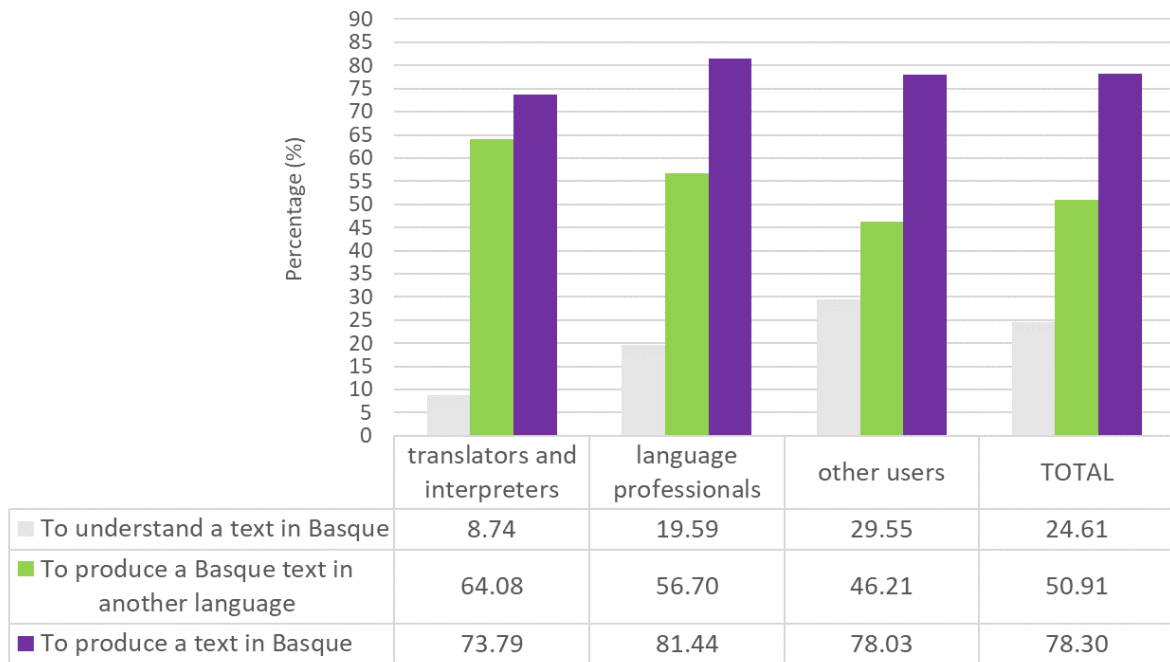


Figure 5: What is your objective when using MT?

It is interesting to see, however, that despite the advanced level of Basque reported by most of the respondents, a very high proportion of respondents use MT to produce a Basque text: 78.30% of the total respondents indicate this. This is true for all language specialization groups. One could argue that this is because respondents consider that the Basque MT of a text written in another language is already a good starting point, and/or that they find it easier to start producing their own texts in another language of preference and produce the Basque version through translation. Respondents also state that they use MT to translate Basque texts to obtain texts in other languages. 64.08% of translators and interpreters report this activity, which decreases to 56.70% in the case of language professionals and lowers to 46.21% for other users.

Regarding the context of use, as with unaided translation, the workplace is once again where MT for the minority language is most frequently used (see Figure 6). Still, we can see that it is also quite regularly used by other users for homework-related tasks. If we compare this with the totality of translation needs observed earlier, we could say that about 30% is very often addressed using MT in the workplace, which is a considerable proportion.

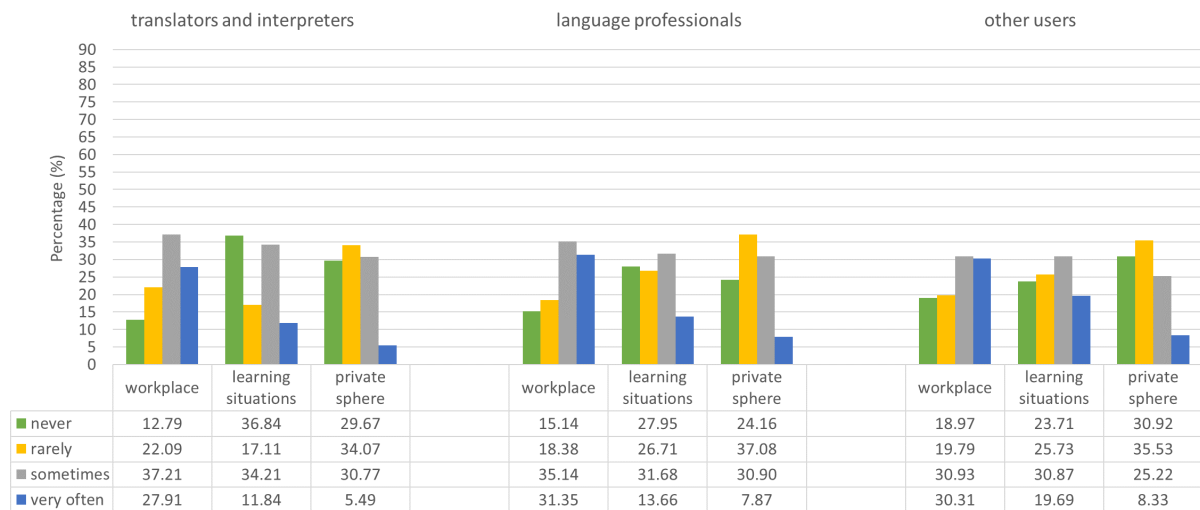


Figure 6: In which contexts do you use MT? Select the frequency.

In terms of text genres, the behaviour of respondents seems to vary slightly in comparison to unassisted translation (see Figure 7). We observe that a higher percentage of other users use MT for specialized texts and essays, which might indicate some awareness of the type of genre that is most successful with the technology. However, we also see a slight increase in its use for literature. Language professionals and translators and interpreters also seem to give preference to informative, more language-neutral texts.

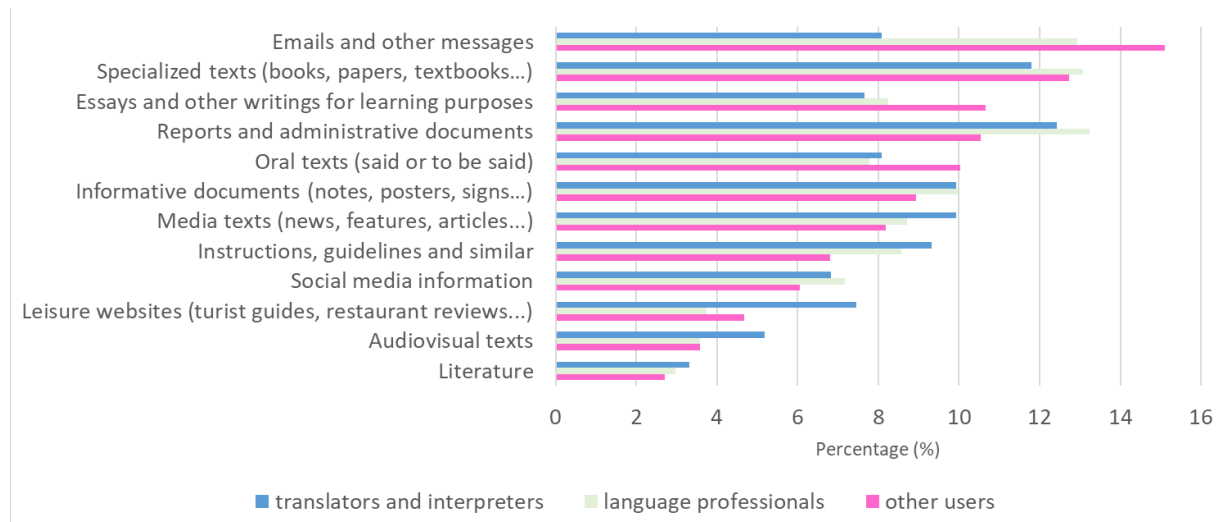


Figure 7: What do you translate [using MT for Basque]? Select all that apply.

We deemed it relevant to find out the units of language the speakers provide the engines to obtain the translations, as quality might depend on it. Current neural MT engines consider context to a very limited degree when constructing the best equivalent. We should also bear in mind that most are trained to translate sentences, which means that they will struggle with smaller units and that they will not consider references or agreements

beyond the sentence structure. Results show that MT is often used as a dictionary, with 27.91% of the total respondents feeding the engines single words (see Figure 8 for the figures per specialization group). Other users approach MT in this manner the most, compared to translators and interpreters, who are the least likely to take this approach. This might reflect their greater awareness of the technology’s internal workings. The language unit that is most frequently provided to the engines is the sentence across all groups, while complete texts are provided by the lowest proportion of speakers.

Yet another relevant issue to ensure good MT practices is what respondents do with the translation suggestions from the MT engines. As can be extracted from the results, the number of respondents who never or rarely use translations unedited is very low, which means that, across all groups, speakers are aware that raw texts are not always adequate. Even other users report fixing the suggestions often (30.11% sometimes and 53.41% very often) (see Figure 9).

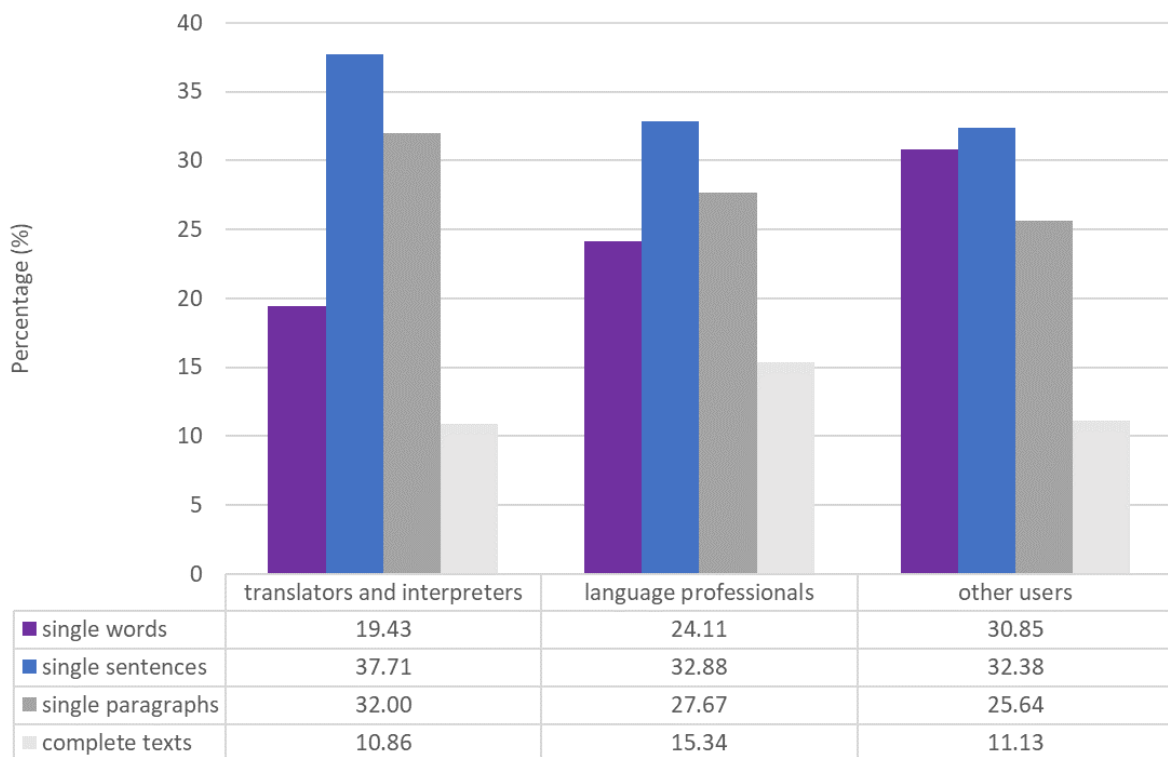


Figure 8: What do you provide the MT engine for translation?

We can view this in a positive light, but there is still work to be done in this area, as we are still uncertain about the effectiveness of the editing process. These results and differences are also relevant from an MT development perspective. Current engines are trained using large parallel corpora. And it is the language and style of those texts that they learn to generate. Most often, in the case of Basque, it is administrative and news-related texts that are available for training. The language used in these training data is not

necessarily the most adequate for translating specialized texts and emails, so we might not be catering to the needs of users.

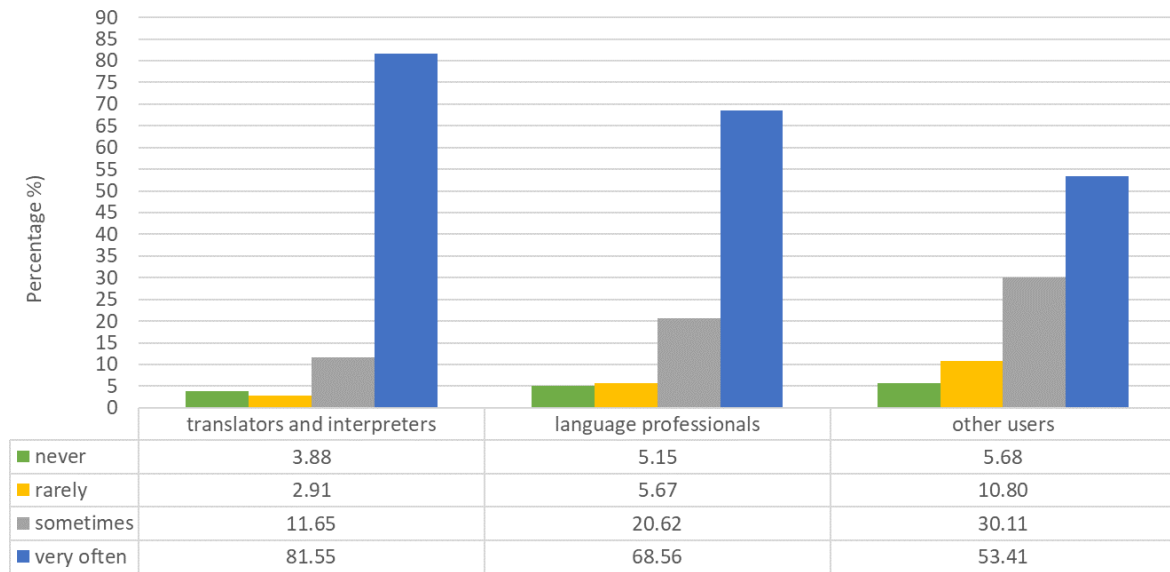


Figure 9: What do you do with the translation proposals generated by MT? How frequently do you adapt them and correct them?

The survey’s results have so far revealed that a large proportion of respondents use MT for Basque. Let us investigate their opinion regarding usefulness (see Figure 10). Most respondents consider the technology useful, either sometimes or very often. We could claim that, in all, 91.39% of the total respondents have a positive view of MT. Language professionals seem to be the ones with the highest level of satisfaction, whereas, expectedly, translators and interpreters are more demanding.

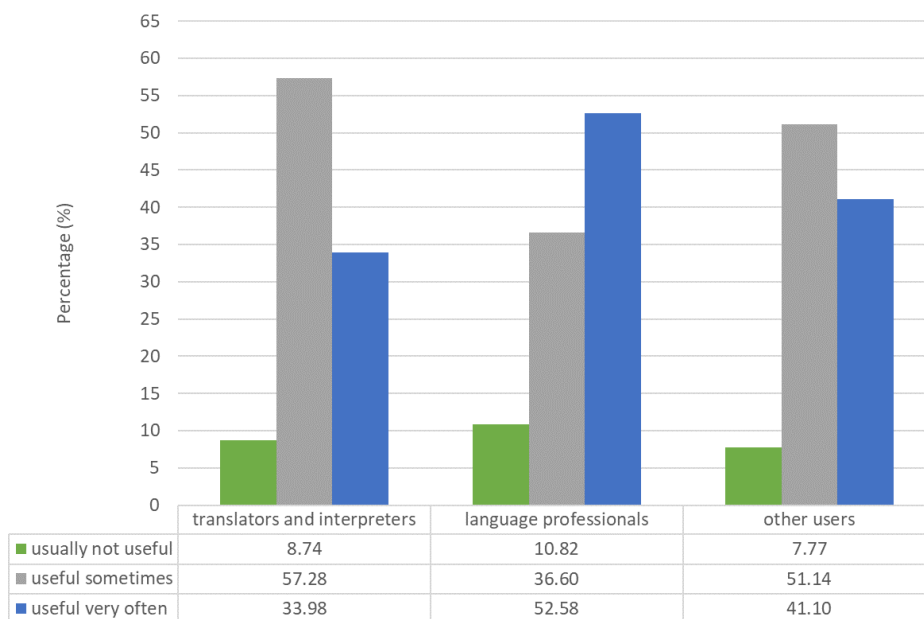


Figure 10: What is your opinion about the quality of Basque MT?

The seemingly positive attitude towards MT for Basque is further confirmed by the responses about the expected MT use in the future (see Figure 11). The great majority of respondents believe that they will use automated translation more often. More than half claim that they are sure about that claim. Among the total respondents, fewer than 5% doubt that their use of MT will increase. The most reluctant seem to be translators and interpreters (4.85%) but only 1.33% of other users share this belief. This is a clear sign that the technology is here to stay and therefore it is essential that we take measures to ensure that it is part of a sustainable revitalization process.

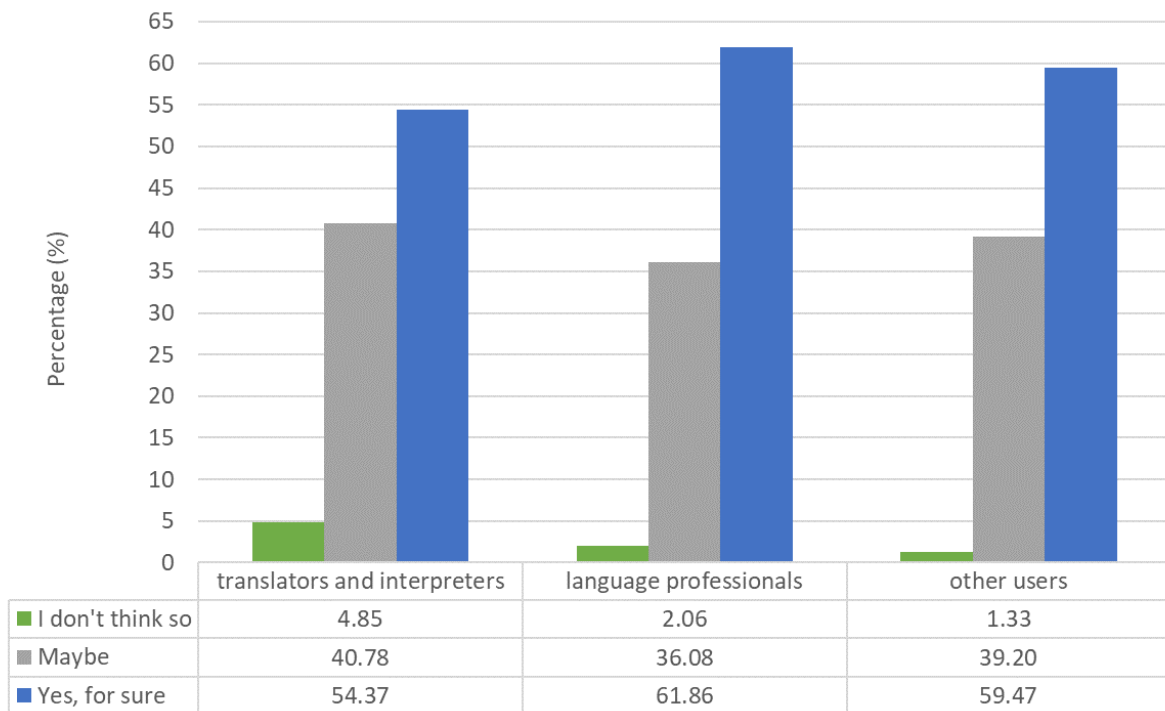


Figure 11: In view of the fact that Basque MT provides better quality output every time, do you expect to use it more in the future?

A final point we deem interesting to discuss in this block is the breadth of dissemination the texts produced with MT have. This is a first indication of the impact the language the MT texts present could have in the development of the target language. According to the data, the texts with a higher level of dissemination are those produced by translators and interpreters, which we expect to be best trained to deal with MT output (see Figure 12). However, we must not overlook the fact that around 50% of respondents in the language professionals and other user groups also affirm that their texts get some exposure. Additionally, for these two groups, 16.46% and 11.66% of respondents claim that the texts are available to a large audience.

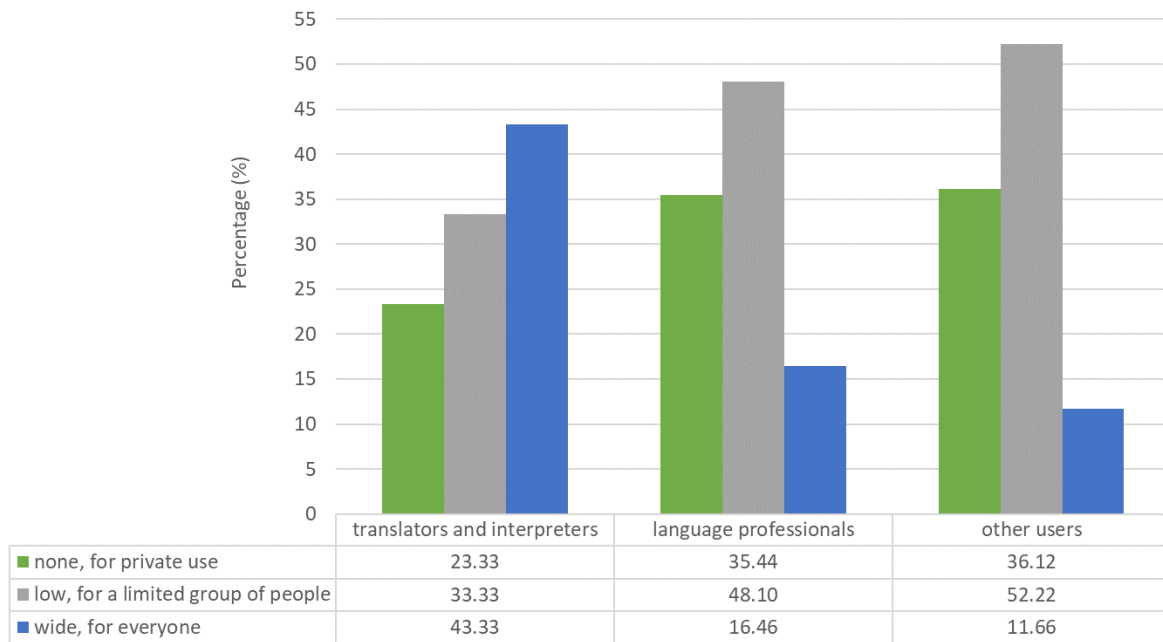


Figure 12: What is the level of dissemination of your translations? You may select more than one option if the answer varies according to the type of translation.

3.6. Results III: impact of machine translation

The third block of the survey consisted of three questions regarding the relation between MT and Basque. While there are many aspects to consider at this stage, to get an initial feel of the respondents’ perceptions, we focused on three points: text production, access, and language quality.

Respondents were asked to state to what extent they agreed to three statements. The first reads as follows: “Seeing that MT quality is increasingly improving; we will write less in Basque every time. Instead, we will tend to write in a major language and have those texts translated automatically into Basque.”

Overall, almost 68.59% of the total respondents do not agree that original production in Basque will decrease due to the ease with which MT will allow users to start, say, in Spanish, and then obtain a good-quality Basque version automatically (see Figure 13). Yet, 31.41% consider that this might be the case. While no major differences emerge from responses by language specialization groups in general, although translators and interpreters appear slightly more pessimistic, a look at the results based on respondents who use MT for Basque and those who do not display an interesting divergence for other users. Other users who believe that Basque production will drop is 13 points higher among those who do not use MT for Basque, which accounts for 40% of respondents in this category.

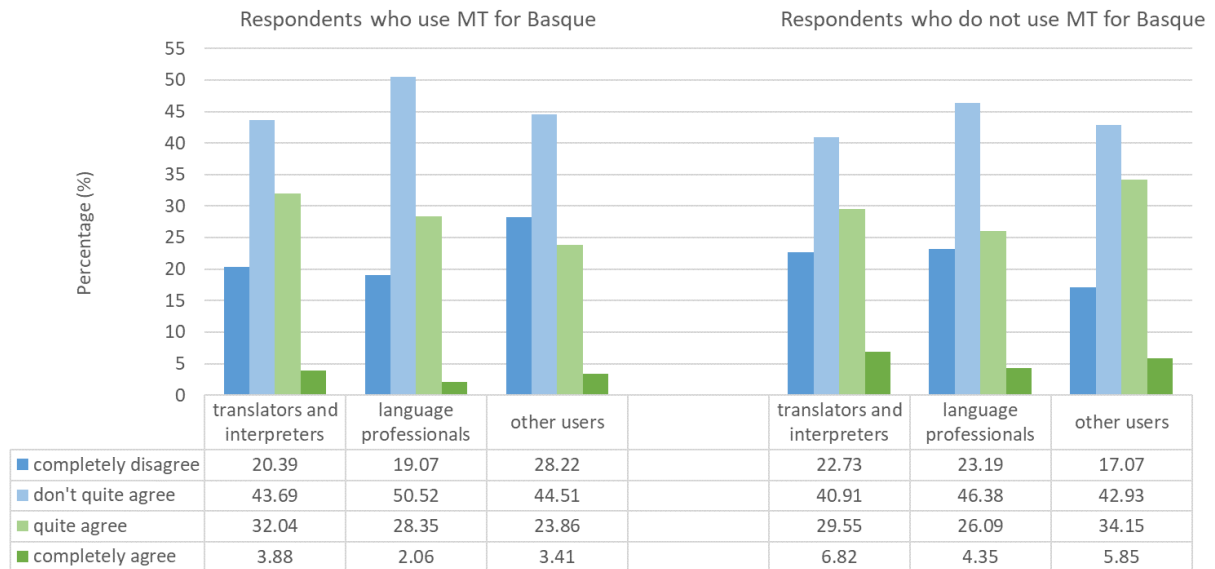
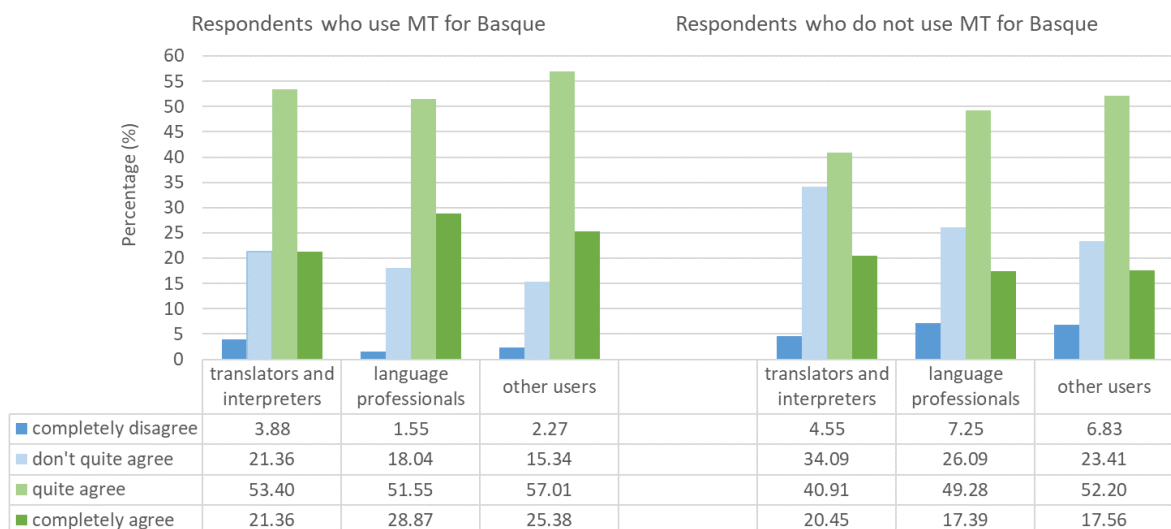


Figure 13: Level of agreement to Statement 1: less and less will be written originally in Basque as people will tend to write in other languages and translate the texts into Basque.

The second statement revolved around the accessibility of Basque texts. The statement read as follows: “Seeing that MT quality is increasingly improving, we will have access to more texts in Basque because more will be translated.”

In general, 77.34% of respondents consider that, indeed, MT will allow for an increased visibility of Basque (see Figure 14). If we consider language specialization groups, we observe that the lower the specialization, the higher the belief that this will be the case (70.75%, 76.81% and 78.85% of translators and interpreters, language professionals and other users agree to the statement). Respondents who currently use MT for Basque agree most, with a difference of almost 15 points across groups.



The final question focused on the quality of the language of the texts translated with the aid of MT engines. Respondents were asked to indicate their level of agreement with the following statement: “If MT is used, the quality of Basque texts risks getting worse.”

If we consider the responses in general, 52.49% disagree with the idea that the use of MT will deteriorate the quality of Basque texts, that is, using the technology will not have any impact on the language we will use and see published (see Figure 15). However, a closer look at the data reveals interesting tendencies. If we consider language specialization groups, we can see that 67.35% of translators and interpreters believe that the quality will drop, in contrast to the considerably lower 42.97% of language professionals and 45.16% of other users. What is also interesting is that, if we consider the responses based on respondents’ use of MT for Basque, the idea that MT could negatively impact the language is more widespread among those who do not use the technology.

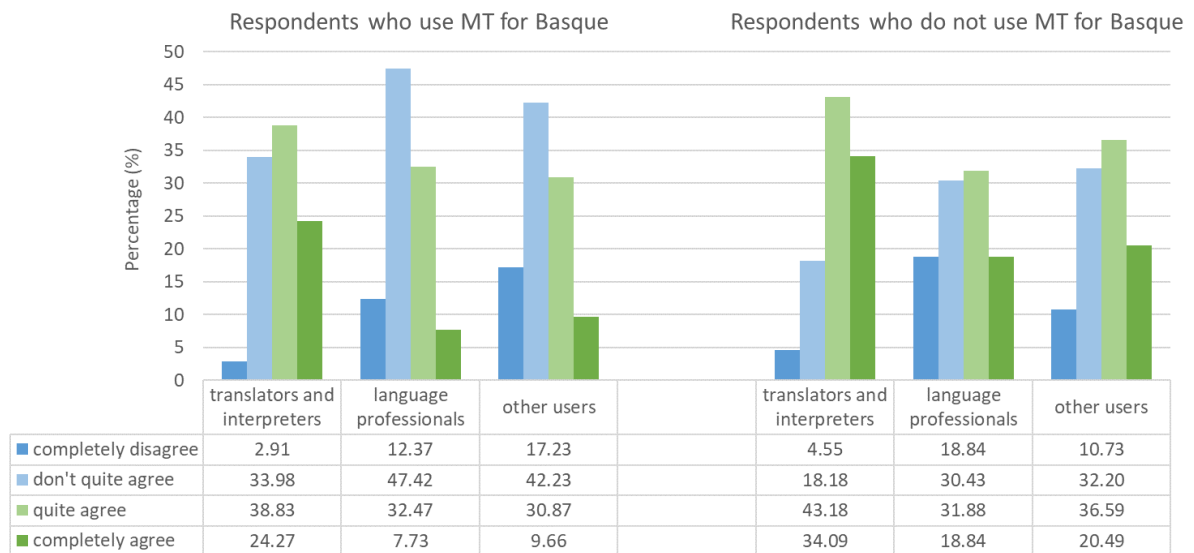


Figure 14: Level of agreement to Statement 3: If MT is used, the quality of Basque texts risks getting worse.

4. Focus group with professional translators

In this section we report on the ideas exchanged in a focus group with professional translators run as part of a series of three online post-editing workshops within EIZIE and which brought together more than 70 specialists (over 20 per workshop). Their experience with MT and post-editing varied from having the technology integrated into their translation memory (TM) environments to not having used the engines for Basque since neural systems emerged and drastically changed the scene for Basque MT.

Given their diverse backgrounds, we got participants working on several translation and post-editing tasks which served as a springboard to discuss

MT in general. To this end, they could choose between administrative, science textbooks and literary excerpts. They performed the tasks in PET (Aziz *et al.* 2012), a tool for post-editing MT output which allowed us to collect several process- and product-oriented data as feedback for the discussions. Participants also had several videos on the technical aspects of MT and more theoretical aspects of post-editing at their disposal throughout the workshops.

In the following paragraphs we summarize the main ideas and opinions that were voiced by participants during the discussions, but let us start with a few words from us, leading the workshop. As Cortés and Jauregi (2019) stated, even when post-editing is becoming a more widespread activity among Basque translators, it is a relatively new practice for this community. As such, we found that it is not always clear what the task involves: is it translation? Is it reviewing? The concept, however obvious it may seem for a seasoned specialist, is not yet clear to many professionals or clients. A few participants reported having received reviewing jobs for which machine translated text was provided. In their opinion, this has a direct impact on the performance of these professionals and the profession itself, as end-product quality, expectations and fees depend on it.

It is also interesting to note that most participants were not familiar with the dynamic nature of quality. While the dichotomy between light and full post-editing is nowadays debatable (Nunziatini and Marg 2020), not all clients expect a quality that rivals human translation. Of course, this raises questions about what it is that should be fixed or not in the MT output. Currently available guidelines from translation associations and research tend to be very general, and often fail to resolve the doubts raised by professionals, who apparently end up making decisions based on time constraints.

Certain reservations were expressed regarding MT output and fixes. Participants stated that many of these edits involve style, both the inherent style of the text but also that of the translator – they would not be able to “make the text their own” – staying too close to the MT proposal and disregarding alternative structures and expressions, and also the restriction of creativity. In general, participants expressed that they would feel more comfortable with jobs requiring high quality, even when less demanding opportunities might be initially tempting. They claimed that they would feel uncomfortable about leaving awkwardly-phrased Basque unedited. Some admitted they may be able to accept it if the translations were only available to a limited, small, specialized group. The level of dissemination seemed to play a role.

An interesting issue was highlighted by participants working on Basque administration. MT has been reported to work well with administrative texts, which include informative discourse, without metaphors or connotations. Also, administrative texts are one of the main training sources of Basque

engines partly because the Basque Institute of Administration releases their TMs. However, as a translator who works for the translation service of the Government of Navarre mentioned, the administrative style of Basque is still developing, and considerable work is still needed to set style patterns and improve communicability, an objective of the current revitalization process. They pointed out that, in this context, translation plays a key role because most administrative texts are first written in Spanish. Therefore, participants wondered whether the field was ready to be disrupted by MT, which if left unattended, might set the main features of the discourse.

Along these lines, some participants expressed concern about the development of Basque in general. A translator and translation lecturer summarized their thoughts: The language, she says, relies heavily on translation, with considerable but proportionally limited original textual production. In this context, MT can act as a double-edged sword. On the one hand, it can promote the translation of Basque texts into other languages so that knowledge produced in the region is disseminated in a timely manner. However, she emphasises that, if we resort to MT for most translations, knowing that the output will be based on training corpora from previous texts, we may risk entering a vicious circle. The linguistic and stylistic resources of Basque could remain undeveloped whereas the new expressions, imagery and resources developed by the hegemonic languages will be reinforced.

Participants working on localization for international language service providers showed a contrasting view of Basque. According to a freelance translator, while clients and LSPs within the Basque community might display a certain degree of awareness and concern about the Basque language and act accordingly, for the majority of multinational companies Basque is just another language. Not just that, she continues, it is a small language among hundreds of big languages that open far larger markets. Some companies are aware that MT for Basque is not yet at the level of that of well-resourced languages, and therefore, apply adequate approaches and rates. However, in her view, with the significant improvement of neural engines, it is unclear how long this “privileged” status will hold.

An in-house translator summarized the view of professionals who currently avail of MT within their translation flows. She stated that MT should not impose a new way to translate, but serve as an additional tool to help translate, integrated within the translator’s workstation. It is by supporting this view that MT is allocated to where it currently belongs. In short, participants claimed that in-house translators prefer to work with MT output rather than starting from scratch as long as they can give preference to TM segments.

Let us conclude with an interesting wish list by in-house translators to improve the interaction between the MT and TM technologies: (1) an estimation of the accuracy of the MT output in terms of closeness to the

original and reliability, (2) a live interaction between the TM glossary and the MT engine, (3) an MT proposal for the unmatched parts of fuzzy matches rather than the complete segment, (4) an MT that considers if not the whole document, at least the near context, and (5) two or three very different alternatives from the MT engine. We must note that several features outlined here are already provided by certain CAT tools, but we are yet to see an environment that fully supports the totality of functionalities suggested by the translators in the focus group.

Lastly, a quick mention of the field of interpreting. As revealed by an in-house interpreter, a few freelancers already work with Basque MT. Participants stated that it allows them to get the gist of the speeches quickly, spot difficult passages and relevant terminology, and that this is key in a context where information for the interventions is provided to the interpreters only a couple of days in advance and often when already in the booth. According to her, interpreters are already familiar with OCR applications but are still waiting for voice recognition to be made widely available. Interestingly, this opinion was expressed with no hint of fear about the future of interpreters.

5. Conclusions

For over 50 years now, Basque has undergone a revitalization process that has seen a successful move towards the normalization of Basque. However, challenges still lay ahead given the changing socioeconomic situation. Within this context, MT emerges as a technology that can assist in addressing language needs but can also jeopardize sustainable development if its impact is not carefully considered. In this work, we aimed to investigate the level of adoption, practices, and perception of MT within the Basque community. We take this as a first step to learn about the current use and expectations so that we can identify the research that will be relevant to take full advantage of the technology in the sustainable revitalization process.

The survey conducted about Basque MT and opinions revealed that translation is a common activity among respondents and, for over 85%, it involves Basque. In this scenario, Basque MT has already reached speakers, 70% of whom use it for many purposes, but most commonly in the workplace. Data shows that a higher awareness of the workings of the engines could be beneficial, as speakers do not seem to be taking full advantage of MT by often inputting inefficient translation units and sometimes leaving proposals unedited. Nonetheless, the overall attitude towards MT seems very positive: 42% consider the output useful very often and a further 48% sometimes. Overall, 60% of respondents are sure that their use of MT will increase in the future and only 2% believe that it will not. What is more, MT is thought to contribute to the visibility of Basque, as 77% agree that more will be translated into the minoritized language thanks to the technology and, according to 68% of the respondents, availing

of MT technology should not result in a decrease in original production in Basque. However, respondents differ in their opinion about the impact MT will have on the quality of the language: while around half believe that it will have no influence, the other half believe that quality will drop. This belief is more marked among translators and interpreters.

As language specialists who have full awareness of the translation process, professional translators and interpreters tend to be more critical towards MT. All in all, however, they seem interested in the technology and ready to either keep using it or to give it an opportunity. Their main concern is related to the impact the technology could have on the evolution of Basque stylistics, which are still developing in several areas. Nonetheless, they seem enthusiastic about seeing a stronger interaction between TM and MT.

The present research shows that Basque MT is already widely used among speakers and will therefore be part of the revitalization process. Now, if we want to ensure the continuation of a sustainable process, research on the impact that MT can have on the target language and literacy initiatives is key.

Acknowledgements

We would like to thank all respondents who voluntarily participated in the survey as well as all the professionals who joined the post-editing workshops and voiced their opinions, in particular, Isabel Etxeberria, Lierni Garmendia, Idoia Gillenea, Ixiar Iza and Alaitz Zabaleta. We also wish to thank the reviewers and editors for their insightful comments.

References

- **Aranberri, Nora** (2020). "With or without you? Effects of using machine translation to write flash fiction in the foreign language." *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*. Lisboa: EAMT, 165–174.
- **Aziz, Wilker, Sheila Castilho and Lucia Specia** (2012). "PET: a Tool for Post-editing and Assessing Machine Translation." In *Proceedings of the 8th International Conference on Language Resources and Evaluation*. Istanbul: LREC, 3982–3987.
- **Baztarrika, Patxi** (2019). "El modelo vasco de revitalización lingüística y desarrollo sostenible". *Linguistic diversity, minority languages and sustainable development*. Bilbao: Servicio Editorial de la UPV/EHU, 239–262.
- **Bowker, Lynne** (2020). "Machine translation literacy instruction for international business students and business English instructors." *Journal of Business & Finance Librarianship* 25(1–2), 25–43.
- **Bowker, Lynne and Jairo Buitrago Ciro** (2019). *Machine translation and global research: Towards improved machine translation literacy in the scholarly community*. Leeds: Emerald Publishing.
- **Cortés, Itziar and Saroi Jauregi** (2019). "Itzultzaile automatikoa: mehatxua ala aukera?". *Senez: itzulpen aldizkaria* 50, 225–230.
- **De Clercq, Orphée, Gert de Sutter, Rudy Loock, Bert Cappelle and Koen**

- Plevoets** (2021). "Uncovering machine translationese using corpus analysis techniques to distinguish between original and machine-translated French". *Translation Quarterly* 101, 21-45.
- **Delorme Benites, Alice** (2021). "Peut-on traduire l'identité? De la différence entre textes traduits et textes originaux en Suisse romande." *Travaux neuchâtelois de linguistique*, (74), 147-166.
 - **Eustat** (2022). "Population and Housing Census. Basque. 2021." Press Release. 27/12/2022. https://en.eustat.eus/elementos/ele0020200/ti_in-2021-624-of-people-residing-in-the-basque-country-had-some-knowledge-of-basque/not0020231_i.html (consulted 10.01.2023).
 - **Eustat** (2016). "VI Encuesta Sociolingüística". https://www.euskara.euskadi.eus/contenidos/informacion/argitalpenak/es_6092/adjuntos/VI%20INK_SOZLG_EAE_Presentacion_publica_20161014.pdf (consulted 10.01.2023).
 - **Gurrutxaga, Antton and Klara Ceberio** (2017). "Basque – a digital language?" In Soria, Claudia, Irene Russo, & Valeria Quochi (eds.), *Reports on digital language diversity in Europe. The Digital Language Diversity Project (DLDP)*. Available at: http://www.dldp.eu/sites/default/files/documents/DLDP_Basque-Report.pdf
 - **Kenny, Dorothy** (ed.) (2022). *Machine translation for everyone: Empowering users in the age of artificial intelligence*. Berlin: Language Science Press.
 - **Marchisio, Kelly, Freitag, Markus and David Grangier** (2022). "On systematic style differences between unsupervised and supervised MT and an application for high-resource machine translation". *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Seattle: Association for Computational Linguistics, 2214-2225.
 - **Moorkens, Joss** (2020). "Comparative satisfaction among freelance and directly-employed Irish-language translators." *Translation & Interpreting*, 12(1), 55-73.
 - **Nunziatini, Mara and Lena Marg** (2020). "Machine Translation Post-Editing Levels: Breaking Away from the Tradition and Delivering a Tailored Service". In *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*. Lisboa: EAMT, 309-318.
 - **Nurminen, Mary and Niko Papula** (2018). "Gist MT users: A snapshot of the use and users of one online MT tool." In *Proceedings of the 21st annual conference of the European Association for machine translation*. Alicante: EAMT, 199-208.
 - **Rossetti, Alessandra, O'Brien, Sharon and Patrick Cadwell** (2020). "Comprehension and trust in crises: investigating the impact of machine translation and post-editing." In *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*. Lisboa: EAMT, 9-18.
 - **Sánchez-Gijón, Pilar and Ramon Piqué Huerta** (2020). "Conseqüències de la traducció automàtica neuronal sobre les llengües d'arribada". *Tradumàtica* 18, 1-10.
 - **Soria, Claudia** (2023). "Decolonizing Minority Language Technology. State of the Internet's Languages Report initiative." *The State of the Internet's Languages Report*, a collaboratively researched and documented initiative, 8 May 2023, 8:00, internetlanguages.org.
 - **Vanmassenhove, Eva, Shterionov, Dimitar and Matthew Gwilliam** (2021). "Machine Translationese: Effects of Algorithmic Bias on Linguistic Complexity in Machine Translation." In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*. Kyiv, Association for Computational Linguistics, 2203-2213.
 - **Vieira, Lucas Nunes** (2020). "Machine translation in the news: A framing analysis of

the written press." *Translation Spaces*, 9(1), 98-122.

- **Vieira, Lucas Nunes, Carol O’Sullivan, Xiaochun Zhang and Minako O’Hagan** (2023). "Machine translation in society: insights from UK users." *Language Resources & Evaluation* 57, 893–914. <https://doi.org/10.1007/s10579-022-09589-1>

Biographies

Nora Aranberri is an associate professor in Translation and Interpreting at the University of the Basque Country and a researcher at HiTZ – Basque Center for Language and Technology, focusing on MT evaluation. She is mainly interested in studying MT use by both professional translators and general users. She works with language pairs involving Basque to explore the implications MT can have for low-resource and minority languages.

ORCID: 0000-0003-3719-9167

Email address: nora.aranberri@ehu.eus



Uxoia Iñurrieta is a lecturer and researcher at the GOI institute of the Basque Summer University (UEU). She graduated in Translation and did her PhD at the HiTZ Basque Center for Language Technology, University of the Basque Country (UPV/EHU). Her teaching and research areas include Language Technologies, Linguistics for Language Teaching and Basque. She is an active member of the Association of Basque Translators (EIZIE) and has translated over 30 books of children and young adult literature.

ORCID: 0000-0002-9807-925X

Email address: u.inurrieta@ueu.eus

